

Abstract of the Disclosure

A system encodes videos acquired of a moving object in a scene by multiple fixed cameras. Camera calibration data of each camera are first determined. The camera calibration data of each camera are associated with the corresponding video. A segmentation mask for each frame of each video is determined. The segmentation mask identifies only foreground pixels in the frame associated with the object. A shape encoder then encodes the segmentation masks, a position encoder encodes a position of each pixel, and a color encoder encodes a color of each pixel. The encoded data can be combined into a single bitstream and transferred to a decoder. At the decoder, the bitstream is decoded to an output video having an arbitrary user selected viewpoint. A dynamic 3D point model defines a geometry of the moving object. Splat sizes and surface normals used during the rendering can be explicitly determined by the encoder, or explicitly by the decoder.